Grant Messner
Section 2

Bi 1 HW 4

1a) Each parent has 1 red eye and 1 white eye allele (since each are orange-eyed).

This means that $P_{red} = P_{white} = .5$ for each draw. Thus, we can calculate:

$f_{rr} = P_{red} * P_{red} = .5 * .5 =. \mathbf{25}$

$f_{orange} = f_{rw} + f_{wr} = P_{red} * P_{white} + P_{white} * P_{red} = .5 * .5 \ + \ .5 * .5 =. \mathbf{5}$

$f_{ww} = P_{white} * P_{white} = .5 * .5 =. \mathbf{25}$

1b)

$$f_r = \frac{2N_{red} + N_{orange}}{2N_{tot}} = \frac{\boldsymbol{2N_{red} + N_{orange}}}{\mathbf{32}}$$

$$f_w = \frac{2N_{white} + N_{orange}}{2N_{tot}} = \frac{\boldsymbol{2N_{white} + N_{orange}}}{\mathbf{32}}$$

1c – g) see Jupyter notebook

2a)

$N_1(t) = N_1(0)e^{m_1 t}$

$N_2(t) = N_2(0)e^{m_2 t}$

$N_{tot}(t) = N_1(t) + N_2(t) = \boldsymbol{N_1(0)e^{m_1 t} + N_2(0)e^{m_2 t}}$

2b) $p(t) = \frac{N_1(t)}{N_1(t)+N_2(t)} = \frac{N_1(0)e^{m_1 t}}{N_1(0)e^{m_1 t}+N_2(0)e^{m_2 t}} = \frac{1}{1+\frac{N_2(0)}{N_1(0)}e^{(m_2-m_1)t}} = \frac{1}{1+\frac{N_2(0)}{N_1(0)}e^{m_1(1-s)t-m_1 t}}$

$$\boldsymbol{p(t) = \frac{1}{1 + \frac{N_2(0)}{N_1(0)} e^{-m_1 st}}}$$

$$q(t) = 1 - p(t) = 1 - \frac{1}{1 + \frac{N_2(0)}{N_1(0)} e^{-m_1 st}} = \frac{1 + \frac{N_2(0)}{N_1(0)} e^{-m_1 st}}{1 + \frac{N_2(0)}{N_1(0)} e^{-m_1 st}} - \frac{1}{1 + \frac{N_2(0)}{N_1(0)} e^{-m_1 st}}$$

$$q(t) = \frac{\frac{N_2(0)}{N_1(0)} e^{-m_1 st}}{1 + \frac{N_2(0)}{N_1(0)} e^{-m_1 st}} = \boldsymbol{\frac{1}{1 + \frac{N_1(0)}{N_2(0)} e^{m_1 st}}}$$

2c) see Jupyter notebook

2e) If we consider there to be only a finite number of potential beneficial mutation sites in the bacteria genome, then the model we observe actually makes a lot of sense. These mutations occur at random, and at first, many of the randomly chosen mutations occur in sites that are beneficial, because there are many beneficial mutation sites that are still "open." However, over time the number of "open" beneficial mutation sites decreases as more and more become randomly chosen, and thus the number of open beneficial sites falls. Therefore, the probability of a random mutation occurring in a beneficial site will fall as time goes on, and thus the rate at which the relative fitness of the bacteria strand goes up will start to decrease, which is exactly what we observe in the graph.

2f)

- Lower bound frequency for standard detection is .01 (with a 1% error rate)
- If we apply this standard detection 50 times to a genome, then we would expect to detect $\frac{.01}{50} = .0002$ as a lower bound in a perfect world, but this comes at a higher error rate
- To find this new error rate, consider the "success rate" for the standard detection, which is $1 - error\ rate = 1 - .01 = .99$
- If an allele is to be detected "successfully," it must be detected successfully 50 times in the deep sequencing technique, which has a probability of $.99^{50} = .605$. Thus, the new error rate for the deep sequencing is $1 - success\ rate = 1 - .605 = .395$
- Therefore, we have $.0002$ as the lower bound at 39.5% error, so this corresponds to $.0002 * \left(\frac{39.5}{1}\right) \approx .$ **008 as the lower bound for the new sequencing techniques**